Journal of Multivariate Analysis 101 (2010) 718–732

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Empirical likelihood inference in partially linear single-index models for longitudinal data

Gaorong Li [a], Lixing Zhu [b,*], Liugen Xue [a], Sanying Feng [c]

[a] College of Applied Sciences, Beijing University of Technology, Beijing 100124, China
[b] Department of Mathematics, Hong Kong Baptist University, Hong Kong, China
[c] College of Mathematics and Science, Luoyang Normal University, Luoyang 471022, China

ARTICLE INFO

Article history:
Received 18 July 2008
Available online 28 August 2009

AMS 2000 subject classifications:
primary 62J05
62G15
secondary 62G20

Keywords:
Longitudinal data
Partially linear single-index model
Empirical likelihood
Confidence region
Bias correction

ABSTRACT

The empirical likelihood method is especially useful for constructing confidence intervals or regions of parameters of interest. Yet, the technique cannot be directly applied to partially linear single-index models for longitudinal data due to the within-subject correlation. In this paper, a bias-corrected block empirical likelihood (BCBEL) method is suggested to study the models by accounting for the within-subject correlation. BCBEL shares some desired features: unlike any normal approximation based method for confidence region, the estimation of parameters with the iterative algorithm is avoided and a consistent estimator of the asymptotic covariance matrix is not needed. Because of bias correction, the BCBEL ratio is asymptotically chi-squared, and hence it can be directly used to construct confidence regions of the parameters without any extra Monte Carlo approximation that is needed when bias correction is not applied. The proposed method can naturally be applied to deal with pure single-index models and partially linear models for longitudinal data. Some simulation studies are carried out and an example in epidemiology is given for illustration.

1. Introduction

Longitudinal data modeling is a statistical method often used in the experiments that are designed such that responses on the same experimental units are observed at each repetition. Experiments of this type have extensive applications in many fields, including epidemiology, econometrics, medicine, life and social sciences. Let $\{(Y_{ij}, X_{ij}, Z_{ij})_{1\le i\le n, 1\le j\le m_i}\}$ be the jth repeated observation for the ith subject or experimental unit, where Y_{ij} is the response variable that is associated with the vector of explanatory variables $(X_{ij}, Z_{ij}) \in R^p \times R^q$. Throughout this paper we assume that n increases to push up the total sample size $N = \sum_{i=1}^n m_i$, while $\{m_i\}$ are the bounded sequences of positive integers. This means that n and N have the same order. The partially linear single-index model for longitudinal data has the form

$$Y_{ij} = g(X_{ij}^{\mathrm{T}}\beta) + Z_{ij}^{\mathrm{T}}\theta + e_{ij}, \quad i = 1, \ldots, n,\ j = 1, \ldots, m_i, \tag{1.1}$$

where (β, θ) is an unknown vector in $R^p \times R^q$ with $\|\beta\| = 1$ (where $\|\cdot\|$ denotes the Euclidean norm), $g(\cdot)$ is an unknown univariate link function. Denote by $e_i = (e_{i1}, e_{i2}, \ldots, e_{im_i})^{\mathrm{T}}$ the random error vector of the ith subject, and $\{e_i, i = 1, \ldots, n\}$ are mutually independent with $E(e_i|X_i, Z_i) = 0$ and the positive definite covariance matrix Σ_i, i.e., $\mathrm{Var}(e_i) = \Sigma_i$. The constraint $\|\beta\| = 1$ is for the identifiability of β because $g(\cdot)$ is unknown and only the orientation of β is identifiable.

* Corresponding author.
E-mail addresses: ligaorong@gmail.com (G. Li), lzhu@hkbu.edu.hk (L. Zhu), lgxue@bjut.edu.cn (L. Xue), fsy5801@sina.com (S. Feng).

0047-259X/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2009.08.006

Model (1.1) is flexible enough to cover many important statistical models. For example, when $\theta = 0$ or, equivalently, there are no predictors Z_{ij}, model (1.1) is a longitudinal single-index model with an unknown link function. The appeal of the model is that by focusing on an index $X_{ij}^{\mathrm{T}}\beta$, the so-called "curse of dimensionality" in fitting multivariate nonparametric regression functions is avoided. Chiou and Müller [1] introduced a flexible marginal modeling approach and proposed the estimated estimating equations (EEE) method to estimate the index parameter vector β. When $p = 1$ and $\beta = 1$, model (1.1) becomes the longitudinal partially linear model, which has been investigated in [2–11] and the references therein. When $m_i = 1$, model (1.1) is reduced to the non-longitudinal partially linear single-index model. Treatments of the non-longitudinal partially linear single-index model include [12–16], among others.

In this paper, we use empirical likelihood (EL) introduced by Owen [17,18] to investigate the partially linear single-index model (1.1) for longitudinal data. It is well known that EL is a widely used statistical method and has been applied to many statistical models. In this area, [19] is a comprehensive reference. For the non-longitudinal partially linear single-index model, Zhu and Xue [14] proposed a bias-corrected empirical likelihood method to construct the confidence regions of parameters of interest. Compared with the existing least squares method, the bias-corrected empirical likelihood ratio shares some of the desired features. In particular, it can avoid estimating the parameters (β, θ) by using the iterative algorithm (see [12]) and their complicated asymptotic covariance matrix.

For longitudinal data, we note that due to the within-subject correlation, empirical likelihood needs to be adjusted so that it can be efficiently used. One of approaches is to apply it to obtain quasi-score from every subject rather than every observation. To obtain the most efficient confidence regions of (β, θ), we propose a block empirical log-likelihood ratio function for the parameters of interest to accommodate the correlation inherent in longitudinal data by introducing the working correlation matrix. Unlike the usual empirical likelihood method, the proposed block empirical likelihood procedure takes the observations of the same subject into account as a whole as is done in [8]. The block empirical likelihood procedure is also different from the moving block empirical likelihood procedure of Kitamura [20]. Kitamura's method is used to deal with time series data, while our procedure is used to deal with longitudinal data. You et al. [8] considered the block empirical likelihood for partially linear model, and Xue and Zhu [21] also used a similar approach for varying coefficient model, but they did not consider the within-subject correlation to further obtain the efficient confidence regions. Li, Tian and Xue [11] showed that ignoring this within-subject correlation leads to a loss of efficiency for empirical likelihood applications. In this paper, we show that correct specification of the correlation structure can result in an asymptotically efficient confidence regions. However, for model (1.1), even when the above approach is applied, the limit of the empirical likelihood ratio is no longer chi-square variable while a weighted sum of chi-square variables with unknown weights. It is not convenient to be used unless some Monte Carlo approximation to its distribution or estimated transformation is employed. The main cause to make this problem happened is the use of plug-in estimation for nuisance parameter/function, which makes a non-negligible bias. Thus, to reduce the bias so that the limit is chi-square, we will propose a centerization approach to the weighted residuals of our model such that the bias does not destroy the Wilks phenomenon, even the plug-in estimations of the function $g(\cdot)$ and its derivative $g'(\cdot)$ are involved. We then call it bias-corrected block empirical likelihood (BCBEL). We can prove that the BCBEL ratio is asymptotically of standard chi-square distribution. Also to achieve this, the optimal bandwidth for estimating $g(\cdot)$ is enough in both the estimators of $g(\cdot)$ and $g'(\cdot)$ although the estimator of $g'(\cdot)$ is of slower convergence rate.

The paper is organized as follows. Section 2 defines BCBEL and obtains the standard chi-square as its asymptotic distribution. Then the results can be used to construct the confidence regions for (β, θ). The pure single-index model and the partially linear model for longitudinal data, as the special examples, are also considered. In Section 3, we present the results from simulation studies and a comparison with profile least squares based normal approximation in terms of coverage accuracy and average areas (lengths) of confidence regions (or intervals). Technical proofs are relegated to the Appendix.

2. Methodology and main results

2.1. BCBEL

Assume that the recorded data $\{(Y_{ij}, X_{ij}, Z_{ij}), i = 1, \ldots, n, j = 1, \ldots, m_i\}$ are generated from model (1.1). The primary interest is to construct the confidence regions of (β, θ). Since $\|\beta\| = 1$ means that the true value of β is the boundary point on the unit sphere, $g(X_{ij}^{\mathrm{T}}\beta)$ does not have derivative at the point β. However, we must use the derivative of $g(X_{ij}^{\mathrm{T}}\beta)$ on β when constructing the empirical likelihood ratio. To solve this problem, we suggest the popularly used "delete-one-component" method (see [13,14]). The detail is as follows. Let $\beta = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ and $\beta^{(r)} = (\beta_1, \ldots, \beta_{r-1}, \beta_{r+1}, \ldots, \beta_p)^{\mathrm{T}}$ be a $p - 1$ dimensional parameters vector deleting the rth component β_r. Without loss of generality, we may assume that the true vector β has a positive component β_r, otherwise, consider $-\beta$ or $\beta_r = -(1 - \|\beta^{(r)}\|^2)^{1/2}$. Then, we can write

$$\beta = \beta(\beta^{(r)}) = (\beta_1, \ldots, \beta_{r-1}, (1 - \|\beta^{(r)}\|^2)^{1/2}, \beta_{r+1}, \ldots, \beta_p)^{\mathrm{T}}.$$

The true parameter $\beta^{(r)}$ satisfies the constraint $\|\beta^{(r)}\| < 1$. Thus, β is infinitely differentiable in a neighborhood of the true parameter $\beta^{(r)}$, the Jacobian matrix is

$$J_{\beta^{(r)}} = (\gamma_1, \ldots, \gamma_p)^{\mathrm{T}},$$

where γ_s $(1 \le s \le p, s \ne r)$ is a $(p-1)$-dimensional unit vector with sth component 1, and $\gamma_r = -(1 - \|\beta^{(r)}\|^2)^{-1/2}\beta^{(r)}$. Since β can be determined by $\beta^{(r)}$, we only need to consider the confidence regions of $(\beta^{(r)}, \theta)$.

To construct the empirical log-likelihood ratio function, we first introduce the following matrix notations. Let

$$\begin{aligned}
&X_i = (X_{i1}, X_{i2}, \dots, X_{im_i})^{\mathrm{T}}, \qquad Z_i = (Z_{i1}, Z_{i2}, \dots, Z_{im_i})^{\mathrm{T}},\\
&Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{im_i})^{\mathrm{T}}, \qquad G(X_i\beta) = (g(X_{i1}^{\mathrm{T}}\beta), g(X_{i2}^{\mathrm{T}}\beta), \dots, g(X_{im_i}^{\mathrm{T}}\beta))^{\mathrm{T}},\\
&G'_{\Delta}(X_i\beta) = \mathrm{diag}(g'(X_{i1}^{\mathrm{T}}\beta), g'(X_{i2}^{\mathrm{T}}\beta), \dots, g'(X_{im_i}^{\mathrm{T}}\beta)),
\end{aligned}$$

where the subscript Δ denotes diagonal matrix. Based on the idea of GEE [22], we introduce an auxiliary random vector in terms of the independence of different subjects

$$\eta_i(\beta^{(r)}, \theta) = \Lambda_i V_i^{-1}[Y_i - G(X_i\beta) - Z_i\theta], \tag{2.1}$$

where V_i is an invertible working correlation matrix, possibly depending on a parameter vector τ, which can be estimating by using the method of moments [22]. In (2.1), $\Lambda_i = (G'_{\Delta}(X_i\beta)X_iJ_{\beta^{(r)}}, Z_i)^{\mathrm{T}}$, then

$$G'_{\Delta}(X_i\beta)X_iJ_{\beta^{(r)}} = (g'(X_{i1}^{\mathrm{T}}\beta)X_{i1}^{\mathrm{T}}J_{\beta^{(r)}}, \dots, g'(X_{im_i}^{\mathrm{T}}\beta)X_{im_i}^{\mathrm{T}}J_{\beta^{(r)}})^{\mathrm{T}},$$

where $g'(\cdot)$ is the derivative of $g(\cdot)$ with respect to $\beta^{(r)}$. Note that $E(\eta_i(\beta^{(r)}, \theta)) = 0$ if $(\beta^{(r)}, \theta)$ is the true parameter. Whereas, when $E(\eta_i(\beta^{(r)}, \theta)) = 0$, we can construct an estimation equation $\sum_{i=1}^n \eta_i(\beta^{(r)}, \theta) = 0$. If we assume that $g(\cdot)$ is known, then the solution of the equation is just the generalized least squares estimator of $(\beta^{(r)}, \theta)$. Therefore, by Owen [17], this can be done to accommodate the independence within the observations from different subjects by using the empirical likelihood. A natural block empirical log-likelihood ratio is

$$l(\beta^{(r)}, \theta) = -2\max\left\{\sum_{i=1}^n \log(np_i)\,\middle|\, p_i \ge 0,\ \sum_{i=1}^n p_i = 1,\ \sum_{i=1}^n p_i\eta_i(\beta^{(r)}, \theta) = 0\right\}. \tag{2.2}$$

To use it, we need plug-in estimation for two unknown functions $g(\cdot)$ and $g'(\cdot)$. Consider a local linear smoother (see, [23]). For any (β, θ), the estimators of $g(\cdot)$ and $g'(\cdot)$ are defined by entirely ignoring the within-subject correlation in the following. Lin and Carroll [24] showed that, when standard kernel methods are used, correctly specifying the correlation matrix in fact will result in an asymptotically less efficient estimator for nonparametric part. First we find (a, b) to minimize

$$\sum_{i=1}^n \sum_{j=1}^{m_i} (Y_{ij} - Z_{ij}^{\mathrm{T}}\theta - a - b(X_{ij}^{\mathrm{T}}\beta - t))^2 K_h(X_{ij}^{\mathrm{T}}\beta - t), \tag{2.3}$$

where $K_h(\cdot) = h^{-1}K(\cdot/h)$, $K(\cdot)$ is a kernel function and $h = h_n$ is a sequence of positive numbers tending to zero, called the bandwidth. Let $(\hat{a}, \hat{b})$ be the solution to the weighted least squares problem (2.3). Then, we define the estimators $\hat{g}(t; \beta, \theta) = \hat{a}$ and $\hat{g}'(t; \beta, \theta) = \hat{b}$. Simple calculation yields

$$\hat{g}(t; \beta, \theta) = \sum_{i=1}^n \sum_{j=1}^{m_i} W_{nij}(t; \beta)(Y_{ij} - Z_{ij}^{\mathrm{T}}\theta), \tag{2.4}$$

and

$$\hat{g}'(t; \beta, \theta) = \sum_{i=1}^n \sum_{j=1}^{m_i} \widetilde{W}_{nij}(t; \beta)(Y_{ij} - Z_{ij}^{\mathrm{T}}\theta), \tag{2.5}$$

where

$$W_{nij}(t; \beta) = \frac{N^{-1}K_h(X_{ij}^{\mathrm{T}}\beta - t)[S_{n,2}(t; \beta) - (X_{ij}^{\mathrm{T}}\beta - t)S_{n,1}(t; \beta)]}{S_{n,0}(t; \beta)S_{n,2}(t; \beta) - S_{n,1}^2(t; \beta)}, \tag{2.6}$$

$$\widetilde{W}_{nij}(t; \beta) = \frac{N^{-1}K_h(X_{ij}^{\mathrm{T}}\beta - t)[(X_{ij}^{\mathrm{T}}\beta - t)S_{n,0}(t; \beta) - S_{n,1}(t; \beta)]}{S_{n,0}(t; \beta)S_{n,2}(t; \beta) - S_{n,1}^2(t; \beta)}, \tag{2.7}$$

and

$$S_{n,l}(t; \beta) = \frac{1}{N}\sum_{i=1}^n \sum_{j=1}^{m_i} K_h(X_{ij}^{\mathrm{T}}\beta - t)(X_{ij}^{\mathrm{T}}\beta - t)^l, \quad l = 0, 1, 2.$$

The estimates $\hat{g}(X_{ij}^{\mathrm{T}}\beta; \beta, \theta)$ and $\hat{g}'(X_{ij}^{\mathrm{T}}\beta; \beta, \theta)$ are of slower than $\sqrt{n}$-rates of convergence even when (β, θ) is known. When they are plugged in (2.1) and (2.2), an estimated auxiliary random vector and an estimated block empirical log-likelihood ratio, say $\widetilde{\eta}_i(\beta^{(r)}, \theta)$ and $\widetilde{l}(\beta^{(r)}, \theta)$, can be defined, respectively. Xue and Zhu [13,25] showed that such a natural replacement will result in $\widetilde{l}(\beta^{(r)}, \theta)$ being asymptotically a weighted sum of independent chi-squared variables, each with one degree of freedom and an unknown weight. To construct the confidence regions of $(\beta^{(r)}, \theta)$, we need to use Monte Carlo approximation or estimated transformation to obtain a tractable limiting distribution. This will decrease the accuracies of the confidence regions because the intensive Monte Carlo simulations are required.

To see what makes such a problem, we can analyse briefly the structure of $\widetilde{l}(\beta^{(r)}, \theta)$. Note that in $\widetilde{l}(\beta^{(r)}, \theta)$, the estimated $\eta_i(\beta^{(r)}, \theta) = \Lambda_i(\beta^{(r)}, \theta)V_i^{-1}[Y_i - G(X_i\beta; \beta, \theta) - Z_i\theta]$ with estimated $g(\cdot)$ and $g'(\cdot)$ to obtain estimated G and G'_{Δ}, say, $\hat{G}$, and $\hat{G}'_{\Delta}$, and then $\tilde{\Lambda}_i(\beta^{(r)}, \theta)$. The difference between $\tilde{\eta}_i(\beta^{(r)}, \theta) = \tilde{\Lambda}_i(\beta^{(r)}, \theta)V_i^{-1}[Y_i - \hat{G}(X_i\beta; \beta, \theta) - Z_i\theta]$ and $\eta_i(\beta^{(r)}, \theta)$ is $(\tilde{\Lambda}_i(\beta^{(r)}, \theta) - \Lambda_i(\beta^{(r)}, \theta))V_i^{-1}[Y_i - \hat{G}(X_i\beta; \beta, \theta) - Z_i\theta]$ plus $\Lambda_i(\beta^{(r)}, \theta)V_i^{-1}[\hat{G}(X_i\beta; \beta, \theta) - G(X_i\beta; \beta, \theta)]$. We can prove that the first term is negligible, and the second term is however of nonparametric convergence rate. Thus, the main target is to reduce this bias. The following proposal is dealing with this. Centerization of $\Lambda_i(\beta^{(r)}, \theta)$ is particularly useful. We define the following bias-corrected auxiliary random vector

$$\hat{\eta}_i(\beta^{(r)}, \theta) = \hat{\Lambda}_i(\beta^{(r)}, \theta)V_i^{-1}[Y_i - \hat{G}(X_i\beta; \beta, \theta) - Z_i\theta], \tag{2.8}$$

where V_i is an invertible working correlation matrix. When the working correlation matrix $V_i = I$ (identity matrix), it is assumed that the observations within the same cluster are independent, that is, assuming working independence (see [3]; when $V_i = \Sigma_i$ (the true covariance matrix), it means the true within-subject correlation structures for longitudinal data; in practice, the working correlation matrix V_i possibly depends on a parameter vector τ, which can be estimated by using the method of moments [22]. For example, in panel data, V_i can be estimated by $n^{-1}\sum_{i=1}^n \hat{e}_i\hat{e}_i^{\mathrm{T}}$, where $\hat{e}_i = Y_i - \hat{g}(X_i\hat{\beta}_I) - Z_i\hat{\theta}_I$, where $(\hat{\beta}_I, \hat{\theta}_I)$ and $\hat{g}(\cdot)$ can be obtained by using the estimation procedure of [12] from working independence. In (2.8),

$$\hat{\Lambda}_i(\beta^{(r)}, \theta) = [\hat{G}'_{\Delta}(X_i\beta; \beta, \theta)(X_i - \hat{G}_1(X_i\beta; \beta))J_{\beta^{(r)}}, Z_i - \hat{G}_2(X_i\beta; \beta)]^{\mathrm{T}},$$

and

$$\hat{G}(X_i\beta; \beta, \theta) = (\hat{g}(X_{i1}^{\mathrm{T}}\beta; \beta, \theta), \hat{g}(X_{i2}^{\mathrm{T}}\beta; \beta, \theta), \ldots, \hat{g}(X_{im_i}^{\mathrm{T}}\beta; \beta, \theta))^{\mathrm{T}},$$
$$\hat{G}_l(X_i\beta; \beta) = (\hat{g}_l(X_{i1}^{\mathrm{T}}\beta; \beta), \hat{g}_l(X_{i2}^{\mathrm{T}}\beta; \beta), \ldots, \hat{g}_l(X_{im_i}^{\mathrm{T}}\beta; \beta))^{\mathrm{T}}, \quad l = 1, 2,$$
$$\hat{G}'_{\Delta}(X_i\beta; \beta, \theta) = \mathrm{diag}(\hat{g}'(X_{i1}^{\mathrm{T}}\beta; \beta, \theta), \hat{g}'(X_{i2}^{\mathrm{T}}\beta; \beta, \theta), \ldots, \hat{g}'(X_{im_i}^{\mathrm{T}}\beta; \beta, \theta)).$$

Thus, we obtain a centerized X and Z conditionally on $X_{i1}^{\mathrm{T}}\beta$. This centerization makes asymptotically the conditional uncorrelation between $\hat{\Lambda}_i(\beta^{(r)}, \theta)$ and $[\hat{G}(X_i\beta; \beta, \theta) - G(X_i\beta; \beta, \theta)]$. This uncorrelation plays a key to obtain a much faster convergence rate of this remainder than uncenterized one. Note that

$$\begin{aligned}&\hat{G}'_{\Delta}(X_i\beta; \beta, \theta)(X_i - \hat{G}_1(X_i\beta; \beta))J_{\beta^{(r)}}\\ &\quad = (\hat{g}'(X_{i1}^{\mathrm{T}}\beta; \beta, \theta)(X_{i1} - \hat{g}_1(X_{i1}^{\mathrm{T}}\beta; \beta))^{\mathrm{T}}J_{\beta^{(r)}}, \ldots, \hat{g}'(X_{im_i}^{\mathrm{T}}\beta; \beta, \theta)(X_{im_i} - \hat{g}_1(X_{im_i}^{\mathrm{T}}\beta; \beta))^{\mathrm{T}}J_{\beta^{(r)}})^{\mathrm{T}},\end{aligned}$$

$\hat{g}_1(t; \beta)$ and $\hat{g}_2(t; \beta)$ are the estimators of $g_1(t) = E[X_{ij}|X_{ij}^{\mathrm{T}}\beta = t]$ and $g_2(t) = E[Z_{ij}|X_{ij}^{\mathrm{T}}\beta = t]$ respectively, namely,

$$\hat{g}_1(t; \beta) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} W_{nij}(t; \beta)X_{ij}, \tag{2.9}$$

and

$$\hat{g}_2(t; \beta) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} W_{nij}(t; \beta)Z_{ij}. \tag{2.10}$$

By using the bias correction, $E(\hat{\eta}_i(\beta^{(r)}, \theta)) = o(1)$ if $(\beta^{(r)}, \theta)$ are the true parameters. Therefore, the bias-corrected block empirical log-likelihood ratio is defined as

$$\hat{l}(\beta^{(r)}, \theta) = -2\max\left\{\sum_{i=1}^{n}\log(np_i) \;\middle|\; p_i \geq 0,\ \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} p_i\hat{\eta}_i(\beta^{(r)}, \theta) = 0\right\}. \tag{2.11}$$

By the Lagrange multiplier method, we have

$$p_i = \frac{1}{n}\frac{1}{1 + \lambda^{\mathrm{T}}(\beta^{(r)}, \theta)\hat{\eta}_i(\beta^{(r)}, \theta)}, \quad i = 1, \ldots, n. \tag{2.12}$$

By (2.12), $\hat{l}(\beta^{(r)}, \theta)$ can be represented as

$$\hat{l}(\beta^{(r)}, \theta) = 2\sum_{i=1}^{n}\log\big(1 + \lambda^{\mathrm{T}}(\beta^{(r)}, \theta)\hat{\eta}_i(\beta^{(r)}, \theta)\big), \tag{2.13}$$

where $\lambda = \lambda(\beta^{(r)}, \theta)$ is determined by

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\eta}_i(\beta^{(r)}, \theta)}{1+\lambda^{\mathrm{T}}\hat{\eta}_i(\beta^{(r)}, \theta)} = 0. \tag{2.14}$$

We state the asymptotic behavior of the empirical likelihood ratio in the following theorems.

Theorem 2.1. *Suppose that conditions* C1–C9 *hold in the Appendix. If* $(\beta^{(r)}, \theta)$ *is the true value of the parameter, and the rth component of* β *is a positive number, then*

$$\hat{l}(\beta^{(r)}, \theta) \xrightarrow{L} \chi^2_{p+q-1},$$

where "$\xrightarrow{L}$" stands for convergence in distribution.

Based on Theorem 2.1, $\hat{l}(\beta^{(r)}, \theta)$ can be used to construct confidence regions for $(\beta^{(r)}, \theta)$. For any given $0 < \alpha < 1$, there exists c_α such that $P(\chi^2_{p+q-1} > c_\alpha) = \alpha$, then

$$I_\alpha(\beta^{(r)}, \theta) = \{(\beta^{(r)}, \theta) \in R^{p-1} \times R^q \mid \hat{l}(\beta^{(r)}, \theta) \le c_\alpha, \|\beta^{(r)}\| < 1\}$$

is the confidence regions of $(\beta^{(r)}, \theta)$ with asymptotically correct coverage probability $1 - \alpha$.

Because β is a boundary point on the surface of the unit sphere, the constraint $\|\beta\| = 1$ removes 1 dimension. The actual dimension of the surface of the unit sphere is $p - 1$, and the values of $p - 1$ components of β can completely determine β itself when we assume with no loss of generality the value of the rth component to be positive. Therefore, Theorem 2.1 shows that once we obtain the confidence regions of $(\beta^{(r)}, \theta)$, the confidence regions of (β, θ) can be immediately obtained through the relation $\beta_r = (1 - \|\beta^{(r)}\|^2)^{1/2}$.

2.2. Two special cases: Single-index model and partially linear model

We first consider the pure single-index model for longitudinal data, meaning that there is no linear component in model (1.1). Rewrite the auxiliary random vector

$$\hat{\eta}_i(\beta^{(r)}) = \hat{\Lambda}_i(\beta^{(r)})V_i^{-1}[Y_i - \hat{G}(X_i\beta; \beta)], \tag{2.15}$$

where

$$\hat{\Lambda}_i(\beta^{(r)}) = \left(\hat{g}'(X_{i1}^{\mathrm{T}}\beta; \beta)J^{\mathrm{T}}_{\beta^{(r)}}(X_{i1} - \hat{g}_1(X_{i1}^{\mathrm{T}}\beta)), \ldots, \hat{g}'(X_{im_i}^{\mathrm{T}}\beta; \beta)J^{\mathrm{T}}_{\beta^{(r)}}(X_{im_i} - \hat{g}_1(X_{im_i}^{\mathrm{T}}\beta))\right). \tag{2.16}$$

Let $\hat{l}(\beta^{(r)})$ denote $\hat{l}(\beta^{(r)}, \theta)$ with $\hat{\eta}_i(\beta^{(r)}, \theta)$ replaced by $\hat{\eta}_i(\beta^{(r)})$ in (2.8). We state the following results.

Corollary 2.1. *Suppose that conditions* C1–C9 *hold in the Appendix. If* $\beta^{(r)}$ *is the true value of the parameter, and the rth component of* β *is a positive number, then*

$$\hat{l}(\beta^{(r)}) \xrightarrow{L} \chi^2_{p-1}.$$

Similar to Theorem 2.1, we can determine the confidence region of $\beta^{(r)}$ through χ^2_{p-1}. For any given $0 < \alpha < 1$, there exists c_α such that $P(\chi^2_{p-1} > c_\alpha) = \alpha$, then

$$I_\alpha(\beta^{(r)}) = \{\beta^{(r)} \in R^{p-1} \mid \hat{l}(\beta^{(r)}) \le c_\alpha,\ \|\beta^{(r)}\| < 1\}$$

is the confidence regions of $\beta^{(r)}$ with asymptotically correct coverage probability $1 - \alpha$.

When $\beta = 1$, model (1.1) is reduced to a partially linear model for longitudinal data. In this case, we introduce a random vector

$$\hat{\eta}_i(\theta) = [Z_i - \hat{G}_2(X_i)]^{\mathrm{T}}V_i^{-1}[Y_i - \hat{G}(X_i) - Z_i\theta], \tag{2.17}$$

where $\hat{G}(X_i)$ and $\hat{G}_2(X_i)$ are defined similarly. Let $\hat{l}(\theta)$ denote $\hat{l}(\beta^{(r)}, \theta)$ with $\hat{\eta}_i(\beta^{(r)}, \theta)$ being replaced by $\hat{\eta}_i(\theta)$. Then, we have the following asymptotically chi-squared result.

Corollary 2.2. *Suppose that conditions* C1–C9 *hold in the Appendix. If* θ *is the true value of the parameter, then*

$$\hat{l}(\theta) \xrightarrow{L} \chi^2_q.$$

A confidence region of θ is given by $I_\alpha(\theta) = \{\theta \in R^q \mid \hat{l}(\theta) \le c_\alpha\}$ with $P(\chi^2_q > c_\alpha) = \alpha$.

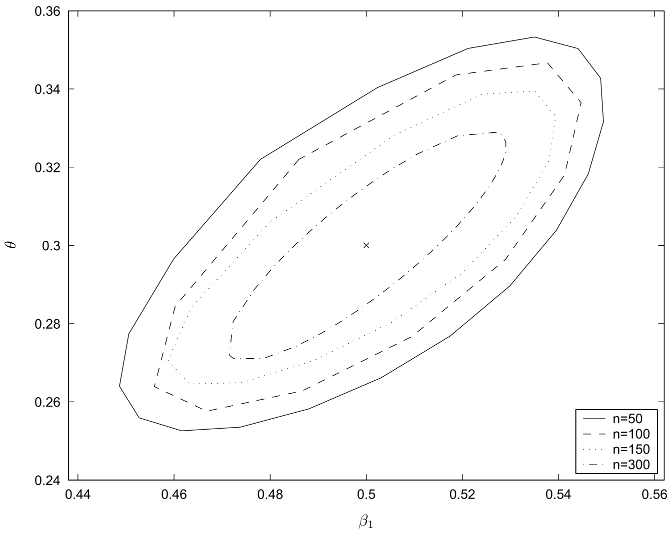

Fig. 1. The 95% confidence regions of (β_1, θ).

Table 1
The coverage probabilities of the confidence regions for (β_1, θ) and (β_2, θ) when the nominal levels are 0.95 and 0.90, respectively.

n	(β_1, θ)		(β_2, θ)	
	95%	90%	95%	90%
50	0.9010	0.8330	0.8940	0.8310
100	0.9240	0.8520	0.9210	0.8570
150	0.9360	0.8820	0.9320	0.8780
300	0.9440	0.8910	0.9420	0.8930

3. Numerical studies

3.1. Simulation studies

In this section, we carry out some simulations to evaluate the finite-sample performance of our proposed method for three models, which are longitudinal partially linear single-index model, longitudinal single-index model and longitudinal partially linear model respectively.

In the following three simulation examples, we generate 1000 datasets, each consisting of $n = 50, 100, 150, 300$ subjects and $m_i \equiv m = 4$ observations per subject. For simplicity, the model error $e_i \sim N(0, \Sigma)$, where $\Sigma = \sigma^2\{(1-\rho)I_m + \rho\mathbf{1}_m\}$ with I_m being the identity matrix and $\mathbf{1}_m$ being a matrix all elements equal to 1, then the within-subject covariance matrix is determined by ρ and σ. The kernel function is taken as the Epanechnikov kernel $K(u) = 0.75(1-u^2)_+$. According to condition C5, the rate of h is between $n^{-1/4}$ and $n^{-1/7}$. Therefore, the bandwidth is selected as $\hat{h} = h_{opt}$ by using the leave-one-subject-out cross validation such that the selected bandwidth $\hat{h}$ satisfies the condition C5. The confidence regions and the coverage probabilities of the confidence regions are computed through the proposed method in Section 2.

Example 1. Consider the partially linear single-index model for longitudinal data

$$Y_{ij} = \sin(X_{ij}^{\mathrm{T}}\beta) + \theta Z_{ij} + e_{ij}, \quad i = 1, \dots, n,\ j = 1, \dots, 4, \tag{3.1}$$

where $\beta = (1/2, \sqrt{3}/2)^{\mathrm{T}}$. The covariate vector X_{ij} is set at $X_{ij} = (X_{1ij}, X_{2ij})^{\mathrm{T}}$, where X_{1ij} and X_{2ij} are generated independently from normal distribution with mean of 0 and variance of 1. $\theta = 0.3$, $Z_{ij} \sim U[0, 1]$. $\sigma = 0.2$, $\rho = 0.6$, Y_{ij} is generated from model (3.1). The numerical results are reported in Table 1, and the confidence regions are given in Figs. 1 and 2.

To avoid estimating the parameter vector (β, θ) by using the iterative algorithm (see [12]), the results of Table 1, Figs. 1 and 2 are obtained by assuming working independence, i.e., $V_i = I$ in (2.8) and (2.11). From Table 1, Figs. 1 and 2, we see that the coverage probabilities based on BCBEL approach the nominal level 0.95 and 0.90 and the confidence regions become smaller as n increases. This implies that the proposed method is valid for the construction of confidence regions based on empirical likelihood.

In the following two examples of the simulations, we consider three different kinds of the working correlation matrices, that is, I (working independence), Σ_i (true correlation matrix) and $\hat{V}_i$ (estimator of unknown case).

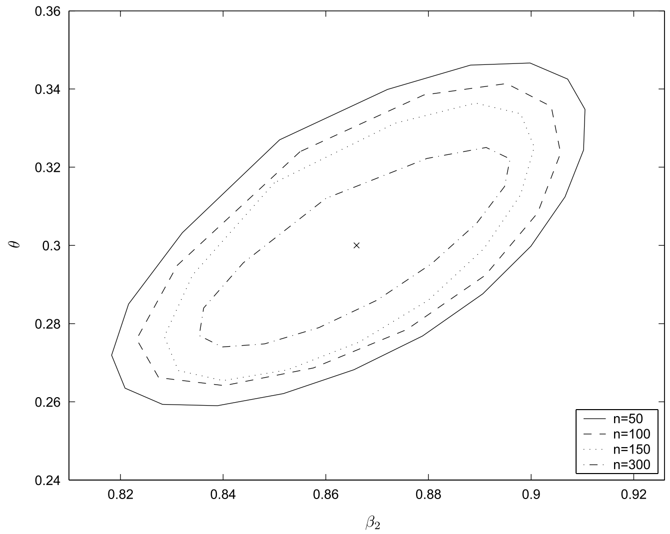

Fig. 2. The 95% confidence regions of (β_2, θ).

Table 2
The CP (AL) of the confidence intervals for β_1 and β_2 based on BCBEL when the nominal levels are 0.95, respectively.

n	CP/(AL)	β_1			β_2		
		I	$\hat{V}_i$	Σ_i	I	$\hat{V}_i$	Σ_i
50		0.8920	0.9050	0.9080	0.8940	0.9070	0.9070
		(0.3253)	(0.2312)	(0.2195)	(0.2876)	(0.2258)	(0.1913)
100		0.9120	0.9180	0.9240	0.9110	0.9220	0.9260
		(0.1972)	(0.1668)	(0.1451)	(0.2010)	(0.1799)	(0.1710)
150		0.9290	0.9310	0.9350	0.9290	0.9330	0.9370
		(0.1553)	(0.1276)	(0.1098)	(0.1481)	(0.1243)	(0.1131)
300		0.9410	0.9460	0.9480	0.9430	0.9450	0.9490
		(0.0987)	(0.0919)	(0.0896)	(0.1002)	(0.0967)	(0.0875)

Example 2. Consider the single-index model for longitudinal data

$$Y_{ij} = \exp(X_{ij}^{\mathrm{T}}\beta) + e_{ij}, \quad i = 1, \ldots, n, j = 1, \ldots, 4, \tag{3.2}$$

where $\beta = (1/2, \sqrt{3}/2)^{\mathrm{T}}$. The covariate vector X_{ij} is set at $X_{ij} = (X_{1ij}, X_{2ij})^{\mathrm{T}}$, where X_{1ij} and X_{2ij} are generated independently from normal distribution with mean of 0 and variance of 1. $\sigma = 0.2$, $\rho = 0.6$, Y_{ij} is generated from model (3.2). In this example, when V_i is unknown, V_i can be estimated by $\hat{V}_i = \frac{1}{n}\sum_{i=1}^{n}\hat{e}_i\hat{e}_i^{\mathrm{T}}$, where $\hat{e}_i = Y_i - \hat{g}(X_i\hat{\beta}_I)$, where $\hat{\beta}_I$ is the profile least squares estimator by assuming working independence. For comparison, two approaches were used in the simulations: the bias-corrected block empirical likelihood (BCBEL) suggested in Section 2, and the profile least squares (PLS). The profile least squares can be obtained in the following. We first obtain the initial estimates for the regression parameter β by sliced inverse regression (see [26]), ignoring the correlation structure within clusters. Then $\hat{g}(\cdot)$ is computed from (2.4) with $\theta = 0$. The solution $\hat{\beta}_V^{(r)}$ $(r = 1, 2)$ of the generalized estimating equations $\sum_{i=1}^{n}\hat{\eta}_i(\beta^{(r)}) = 0$, which is called the profile least squares estimator (PLSE), where $\hat{\eta}_i(\beta^{(r)})$ can be found in (2.15). The comparison was made through coverage probabilities (CP) and average lengths (AL) of confidence intervals. The PLS-based confidence regions were constructed in terms of the asymptotic normal distribution of $\hat{\beta}_V^{(r)}$ with the plug-in estimated asymptotic variance matrix $\hat{\Omega}_{\mathrm{PLS}} = n^{-1}\hat{D}(\hat{\beta}_V^{(r)})^{-1}\hat{B}(\hat{\beta}_V^{(r)})\hat{D}(\hat{\beta}_V^{(r)})^{-1}$, where

$$\hat{D}(\hat{\beta}_V^{(r)}) = \frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_i(\hat{\beta}_V^{(r)})V_i^{-1}\hat{\Lambda}_i^{\mathrm{T}}(\hat{\beta}_V^{(r)}), \qquad \hat{B}(\hat{\beta}_V^{(r)}) = \frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_i(\hat{\beta}_V^{(r)})V_i^{-1}\hat{\Sigma}_i(\hat{\beta}_V)V_i^{-1}\hat{\Lambda}_i^{\mathrm{T}}(\hat{\beta}_V^{(r)}),$$

where $\hat{\Lambda}_i(\hat{\beta}_V^{(r)})$ denotes the estimator of $\Lambda_i(\beta_V^{(r)})$ by plugging $\hat{\beta}_V^{(r)}$ into (2.16). Similar to $\hat{V}_i$, $\hat{\Sigma}_i(\hat{\beta}_V)$ also can be obtained easily.

From Tables 2 and 3, it is easy to see that BCBEL performs much better than PLS in terms of coverage accuracies of the confidence intervals. Tables 2 and 3 also present a comparison for three choices of the working correlation matrices, i.e., Σ_i, $\hat{V}_i$ and working independence matrix I. For the three different kinds of the working correlation matrices, the coverage probabilities obtained by BCBEL and PLS methods tend to 0.95 and the average lengths decrease as n increases. And for

Table 3
The CP (AL) of the confidence intervals for β_1 and β_2 based on PLS when the nominal levels are 0.95, respectively.

n	CP/(AL)	β_1			β_2		
		I	$\hat{V}_i$	Σ_i	I	$\hat{V}_i$	Σ_i
50		0.8940	0.9010	0.9060	0.8980	0.9040	0.9070
		(0.4232)	(0.2531)	(0.2225)	(0.3865)	(0.2488)	(0.2211)
100		0.9090	0.9130	0.9170	0.9070	0.9150	0.9180
		(0.2117)	(0.1812)	(0.1514)	(0.2090)	(0.1831)	(0.1503)
150		0.9250	0.9290	0.9340	0.9250	0.9310	0.9380
		(0.1658)	(0.1396)	(0.1108)	(0.1617)	(0.1354)	(0.1218)
300		0.9380	0.9420	0.9460	0.9410	0.9440	0.9470
		(0.1079)	(0.1065)	(0.0967)	(0.1104)	(0.1029)	(0.0947)

Table 4
The coverage probabilities (CP) of the confidence intervals for θ when the nominal levels are 0.95.

n	BCBEL			PLS		
	I	$\hat{V}_i$	Σ_i	I	$\hat{V}_i$	Σ_i
50	0.9050	0.9120	0.9180	0.8870	0.8990	0.9040
100	0.9190	0.9300	0.9340	0.9130	0.9210	0.9280
150	0.9330	0.9420	0.9450	0.9290	0.9360	0.9440
300	0.9420	0.9470	0.9470	0.9410	0.9430	0.9460

Table 5
The average lengths (AL) of the confidence intervals on θ when the nominal level is 0.95.

n	BCBEL			PLS		
	I	$\hat{V}_i$	Σ_i	I	$\hat{V}_i$	Σ_i
50	0.2971	0.2824	0.2598	0.3772	0.3489	0.3283
100	0.2116	0.1995	0.1893	0.2694	0.2436	0.2210
150	0.1535	0.1467	0.1326	0.1901	0.1836	0.1687
300	0.1127	0.0997	0.0981	0.1225	0.1086	0.1023

the same observation sample, BCBEL and PLS methods perform the best for the true correlation matrix Σ_i, while perform much better for the case of $\hat{V}_i$ than that of I (working independence) in terms of coverage accuracies of the confidence intervals. A reason for the conclusion is that we consider the within-subject correlation structures for Σ_i and $\hat{V}_i$, but ignore the correlation structures for I entirely. Consequently this reduces the accuracies and efficiencies of the confidence regions by ignoring the within-subject correlation structures entirely. The loss of accuracies decreases as n increases, though the accuracies of the confidence regions may possibly be reduced due to mis-specification of the covariance structure.

Example 3. Consider the partially linear model for longitudinal data

$$Y_{ij} = Z_{ij}\theta + \sin(8X_{ij}) + 2 + e_{ij}, \quad i = 1, \ldots, n,\ j = 1, \ldots, 4, \tag{3.3}$$

where $\theta = 2$, Z_{ij} follows uniform $(-0.5, 0.5)$, $X_{ij} \sim N(0, 1)$, $\sigma = 0.8$, $\rho = 0.6$, Y_{ij} is generated from model (3.3). In this example, a comparison between BCBEL and the profile least squares method (PLS) introduced by Fan and Li [5] was made through the coverage probabilities and the average lengths of the confidence intervals. The PLSE is given by

$$\hat{\theta}_V = \left(\sum_{i=1}^{n} \widetilde{Z}_i^{\mathrm{T}} V_i^{-1} \widetilde{Z}_i\right)^{-1} \sum_{i=1}^{n} \widetilde{Z}_i^{\mathrm{T}} V_i^{-1} \widetilde{Y}_i,$$

where $\widetilde{Z}_i = (\widetilde{Z}_{i1}, \widetilde{Z}_{i2}, \ldots, \widetilde{Z}_{im_i})^{\mathrm{T}}$ with $\widetilde{Z}_{ij} = Z_{ij} - \sum_{k=1}^{n}\sum_{l=1}^{m_k} W_{nkl}(X_{ij})Z_{kl}$, and $\widetilde{Y}_i$ can be defined similarly. The plug-in estimated asymptotic variance matrix of $\hat{\theta}_V$ is defined by $\hat{\Omega}_{\mathrm{PLS}} = n^{-1}\hat{D}^{-1}\hat{B}(\hat{\theta}_V)\hat{D}^{-1}$, where

$$\hat{D} = \frac{1}{n}\sum_{i=1}^{n} \widetilde{Z}_i^{\mathrm{T}} V_i^{-1} \widetilde{Z}_i, \qquad \hat{B}(\hat{\theta}_V) = \frac{1}{n}\sum_{i=1}^{n} \widetilde{Z}_i^{\mathrm{T}} V_i^{-1} \hat{\Sigma}_i(\hat{\theta}_V) V_i^{-1} \widetilde{Z}_i.$$

When V_i is unknown, V_i can be estimated by $\hat{V}_i = \frac{1}{n}\sum_{i=1}^{n} \hat{e}_i \hat{e}_i^{\mathrm{T}}$, where $\hat{e}_i = Y_i - Z_i^{\mathrm{T}}\hat{\theta}_I - \hat{g}(X_i)$, where $\hat{\theta}_I = \left(\sum_{i=1}^{n} \widetilde{Z}_i^{\mathrm{T}}\widetilde{Z}_i\right)^{-1} \sum_{i=1}^{n} \widetilde{Z}_i^{\mathrm{T}}\widetilde{Y}_i$, $\hat{g}(x) = \sum_{i=1}^{n}\sum_{j=1}^{m} W_{nij}(x)(Y_{ij} - Z_{ij}^{\mathrm{T}}\hat{\theta}_I)$ is computed from working independence. The numerical results can be found in Tables 4 and 5.

From Tables 4 and 5, it is easy to see that the same conclusions can be obtained similar to Example 2.

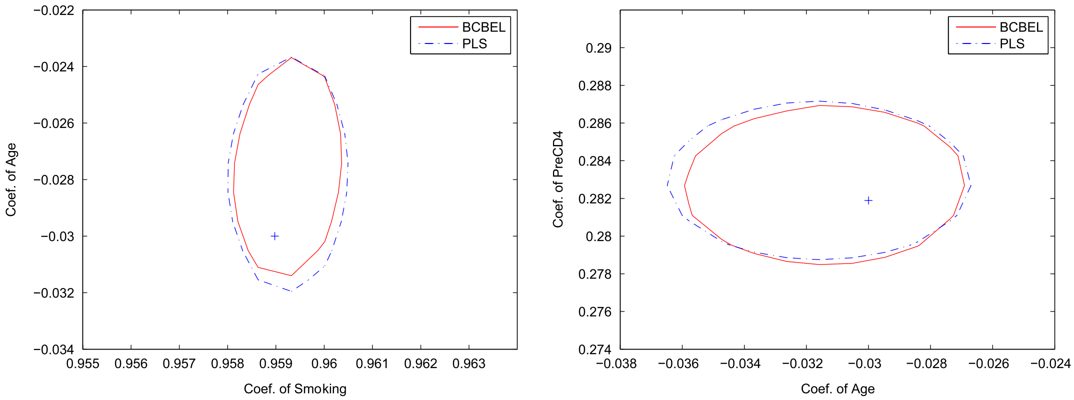

Fig. 3. Application to CD4 data. The 95% confidence regions based on BCBEL and PLS for the coefficients of the smoking effect, the pre-infection CD4 effect and the age effect.

3.2. Application to CD4 data

We now apply the proposed procedure to the CD4 data from the Multi-Center AIDS Cohort Study. The dataset contains the human immunodeficiency virus (HIV) status of 283 homosexual men who were infected with HIV during the follow-up period between 1984 and 1991. All individuals were scheduled to have their measurements made during semiannual visits. Here $t_{ij}, i = 1, \ldots, n, j = 1, \ldots, m_i$, denotes the time length in years between seroconversion and the j-th measurement of the i-th individual after the infection. For various reasons, some individuals missed scheduled visits. Each person was infected randomly during the study. This leads to unequal numbers of repeated measurements m_i and different measurement times t_{ij} across individuals. Details of the design and method of the study have been described by Kaslow et al. [27]. Wu et al. [28], Huang et al. [29], Fan and Zhang [30] and Fan and Li [5] analyzed the same dataset using varying coefficient models. The primary interest was to describe the trend of the mean CD4 percentage depletion over time and to evaluate the effects of cigarette smoking, pre-HIV infection CD4 percentage, and age at infection on the mean CD4 cell percentage after the infection.

In our analysis, the response variable is the CD4 cell percentage of a subject at distinct time points after HIV infection. We take three covariates for this study: X_1, the individual's smoking status, which takes binary values 1 or 0, according to whether a individual is a smoker or nonsmoker; X_2, the CD4 cell percentage level before HIV infection; and X_3, age at HIV infection. We consider the following single-index model:

$$Y(t_{ij}) = g\Big(\beta_1 X_1(t_{ij}) + \beta_2 X_2(t_{ij}) + \beta_3 X_3(t_{ij})\Big) + e(t_{ij}),$$

where β_1, β_2 and β_3 describe the effects for cigarette smoking, pre-infection CD4 percentage and age at HIV infection, respectively, on the post-infection CD4 percentage through a unknown link function $g(\cdot)$. For the identifiability of the model, we require that $\|\beta\| = 1$, where $\beta = (\beta_1, \beta_2, \beta_3)^{\mathrm{T}}$. To construct the confidence regions for the coefficients $(\beta_1, \beta_2, \beta_3)^{\mathrm{T}}$, the Epanechnikov kernel is taken and the "leave-one-subject-out" cross-validation method is employed to select the bandwidth. Using the BCBEL and the normal approximation based on the profile least squares method (PLS), we obtain the 95% confidence regions for the coefficients $(\beta_1, \beta_2, \beta_3)$ of smoking, pre-infection CD4 and age at HIV infection that are shown in Fig. 3. Fig. 3 indicates that, for this dataset, the BCBEL-based confidence region is slightly smaller than the PLS-based region.

Acknowledgments

Gaorong Li's research was supported by China Postdoctoral Science Foundation Funded Project (20080430633) and Shanghai Postdoctoral Scientific Program (08R214121). Lixing Zhu's research was supported by a grant from Research Grants Council of Hong Kong, and a FRG grant from the Hong Kong Baptist University. Liugen Xue's research was supported by the National Natural Science Foundation of China (10871013), the Beijing Natural Science Foundation (1072004), Ph.D. program Foundation of Ministry of Education of China (20070005003) and Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality and the Basic Research. Sanying Feng's research was supported by the Natural Science Research Project of Henan (2008B110009). The authors would like to thank the editor, the associate editor, and the referees for their helpful comments, which helped to improve an earlier version of the article.

Appendix

A.1. Conditions

We now give a set of conditions for the results stated in the theorems.

C1 For any $i = 1, \dots, n, j = 1, \dots, m_i$, the density function $f(t)$ of $X_{ij}^{\mathrm{T}}\beta$ is bounded away from zero on $\mathcal{T}$, and satisfies the Lipschitz condition of order 1 on $\mathcal{T}$, where $\mathcal{T} = \{t = X_{ij}^{\mathrm{T}}\beta : X_{ij} \in A, i = 1, \dots, n, j = 1, \dots, m_i\}$ and A is a compact support set of X_{ij}. And $f(t)$ satisfies $0 < \inf_{t\in\mathcal{T}} f(t) \le \sup_{t\in\mathcal{T}} f(t) < \infty$.

C2 $g(t)$ has two bounded and continuous derivative on $\mathcal{T}$; $g_{1s}(t)$ and $g_{2k}(t)$ satisfy the local Lipschitz condition of order 1, where $g_{1s}(t)$ and $g_{2k}(t)$ are the sth and kth component of $g_1(t)$ and $g_2(t)$ $(1 \le s \le p, 1 \le k \le q)$ respectively.

C3 The kernel $K(u)$ is a bounded and symmetric probability density function, and satisfies

$$\int_{-\infty}^{\infty} u^2 K(u)\mathrm{d}u \ne 0, \qquad \int_{-\infty}^{\infty} |u|^i K(u)\mathrm{d}u < \infty, \quad i = 1, 2, \dots.$$

C4 There exists a positive constant M, such that $\max_{1\le i\le n, 1\le j\le m_i} \sup_{x,z} E(e_{ij}^4 | X_{ij} = x, Z_{ij} = z) \le M < \infty$ and $\max_{1\le i\le n, 1\le j\le m_i} \sup_x E(e_{ij}^4 | X_{ij} = x) \le M < \infty$.

C5 When $n \to \infty$, the bandwidth h satisfies that $h \to 0,\ nh^3 \to \infty,\ nh^8 \to 0$.

C6 There exist two positive constants c_1 and c_2 such that

$$0 < c_1 \le \min_{1\le i\le n} \lambda_{i1} \le \max_{1\le i\le n} \lambda_{im_i} \le c_2 < \infty,$$

where λ_{i1} and λ_{im_i} denote the smallest and largest eigenvalues of Σ_i, respectively.

C7 There exist positive constants c_3 and c_4 such that

$$0 < c_3 \le \min_{1\le i\le n} \lambda'_{i1} \le \max_{1\le i\le n} \lambda'_{im_i} \le c_4 < \infty,$$

where λ'_{i1} and λ'_{im_i} denote the smallest and largest eigenvalues of V_i, respectively.

C8 There exists the positive constant M satisfies, for all i, j, $\sup_{t\in\mathcal{T}} E(\|Z_{ij}\|^2 | X_{ij}^{\mathrm{T}}\beta = t) \le M < \infty$.

C9 $\Omega(\beta^{(r)}, \theta) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n E\{[\Lambda_i - E(\Lambda_i|X_i\beta)]V_i^{-1}(e_i e_i^{\mathrm{T}})V_i^{-1}[\Lambda_i - E(\Lambda_i|X_i\beta)]^{\mathrm{T}}\}$ is a positive matrix, where Λ_i is defined in (2.1).

A.2. Proof of Theorem 2.1

Applying the Taylor expansion to (2.13) and invoking Lemmas A.5–A.7, we can obtain that

$$\hat{l}(\beta^{(r)}, \theta) = 2\sum_{i=1}^n \left[\lambda^{\mathrm{T}}\hat{\eta}_i(\beta^{(r)}, \theta) - \frac{1}{2}\{\lambda^{\mathrm{T}}\hat{\eta}_i(\beta^{(r)}, \theta)\}^2\right] + o_P(1). \tag{A.1}$$

By (2.14), it follows that

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^n \frac{\hat{\eta}_i(\beta^{(r)}, \theta)}{1 + \lambda^{\mathrm{T}}\hat{\eta}_i(\beta^{(r)}, \theta)} \\ &= \frac{1}{n}\sum_{i=1}^n \hat{\eta}_i(\beta^{(r)}, \theta) - \frac{1}{n}\sum_{i=1}^n \hat{\eta}_i(\beta^{(r)}, \theta)\hat{\eta}_i^{\mathrm{T}}(\beta^{(r)}, \theta)\lambda + \frac{1}{n}\sum_{i=1}^n \frac{\hat{\eta}_i(\beta^{(r)}, \theta)(\lambda^{\mathrm{T}}\hat{\eta}_i(\beta^{(r)}, \theta))^2}{1 + \lambda^{\mathrm{T}}\hat{\eta}_i(\beta^{(r)}, \theta)}. \end{aligned}$$

The application of Lemmas A.5–A.7 again yields that

$$\sum_{i=1}^n [\lambda^{\mathrm{T}}\hat{\eta}_i(\beta^{(r)}, \theta)]^2 = \sum_{i=1}^n \lambda^{\mathrm{T}}\hat{\eta}_i(\beta^{(r)}, \theta) + o_P(1), \tag{A.2}$$

$$\lambda = \left[\sum_{i=1}^n \hat{\eta}_i(\beta^{(r)}, \theta)\hat{\eta}_i^{\mathrm{T}}(\beta^{(r)}, \theta)\right]^{-1} \sum_{i=1}^n \hat{\eta}_i(\beta^{(r)}, \theta) + o_P(n^{-1/2}). \tag{A.3}$$

From (A.1)–(A.3), we have

$$\hat{l}(\beta^{(r)}, \theta) = \left[\sum_{i=1}^n \hat{\eta}_i(\beta^{(r)}, \theta)\right]^{\mathrm{T}} \left[\sum_{i=1}^n \hat{\eta}_i(\beta^{(r)}, \theta)\hat{\eta}_i^{\mathrm{T}}(\beta^{(r)}, \theta)\right]^{-1} \left[\sum_{i=1}^n \hat{\eta}_i(\beta^{(r)}, \theta)\right] + o_P(1).$$

By Lemma A.5, we obtain

$$\hat{l}(\beta^{(r)}, \theta) = \left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\eta}_i(\beta^{(r)}, \theta)\right]^{\mathrm{T}} \Omega^{-1}(\beta^{(r)}, \theta) \left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\eta}_i(\beta^{(r)}, \theta)\right] + o_P(1). \tag{A.4}$$

From Lemma A.4, we have

$$\Omega^{-\frac{1}{2}}(\beta^{(r)},\theta)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\eta}_i(\beta^{(r)},\theta) \xrightarrow{L} N(0, I_{p+q-1}), \tag{A.5}$$

where I_{p+q-1} is the $(p+q-1)\times(p+q-1)$ identity matrix. From (A.4) and (A.5), it is easy show that $\hat{l}(\beta^{(r)},\theta)$ is asymptotically chi-squared with $p+q-1$ degrees of freedom. The proof is completed. $\square$

The proofs of Corollaries 2.1 and 2.2 are similar to those used in the proof Theorem 2.1. Therefore, we omit the details.

A.3. Lemmas

For sake of convenience and simplicity, we shall employ c $(0 < c < \infty)$ to denote some constant not depending on n and N but may take difference values at each appearance. The following Lemmas A.1–A.3 generalize Lemmas 1–3 in [14] (the i.i.d. case) to the case of longitudinal data respectively. Because the proofs are very similar to those, we then omit the details here.

Lemma A.1. *Assume that conditions* C1–C3 *hold. If $h = cn^{-\alpha}$, $0 < \alpha < 1/2$, $c > 0$ and $n = O(N)$, then, for $i_1 = 1,\dots,n$, $j_1 = 1,\dots,m_{i_1}$, and any integer $r \geq 2$, we have*

$$E\left\{\left|\sum_{i=1}^{n}\sum_{j=1}^{m_i}W_{nij}(X_{i_1j_1}^{\mathrm{T}}\beta;\beta)\varphi(X_{ij}^{\mathrm{T}}\beta)-\varphi(X_{i_1j_1}^{\mathrm{T}}\beta)\right|^r\right\} = O(h^{2r}), \tag{A.6}$$

$$E\left\{\left|\sum_{i=1}^{n}\sum_{j=1}^{m_i}\widetilde{W}_{nij}(X_{i_1j_1}^{\mathrm{T}}\beta;\beta)g(X_{ij}^{\mathrm{T}}\beta)-g'(X_{i_1j_1}^{\mathrm{T}}\beta)\right|^r\right\} = O(h^{r}), \tag{A.7}$$

where $\varphi(\cdot) = g(\cdot),\ g_{1s}(\cdot)$ and $g_{2k}(\cdot),\ 1 \le s \le p,\ 1 \le k \le q$.

Lemma A.2. *Under the assumptions of Lemma A.1, for any integer $r \geq 2$, we have*

$$\begin{cases} E\{|W_{nij}(X_{ij}^{\mathrm{T}}\beta;\beta)|^r\} = O((nh)^{-r}),\\ E\left\{\displaystyle\sum_{i=1,i_1\neq i}^{n}\sum_{j=1,j_1\neq j}|W_{nij}(X_{i_1j_1}^{\mathrm{T}}\beta;\beta)|^r\right\} = O((nh)^{1-r}),\end{cases}$$

$$\begin{cases} E\{|\widetilde{W}_{nij}(X_{ij}^{\mathrm{T}}\beta;\beta)|^r] = O((nh)^{-r}) + O((n^3h^5)^{-r/2}),\\ E\left\{\displaystyle\sum_{i=1,i_1\neq i}^{n}\sum_{j=1,j_1\neq j}|\widetilde{W}_{nij}(X_{i_1j_1}^{\mathrm{T}}\beta;\beta)|^r\right\} = O(n^{1-r}h^{1-2r}).\end{cases}$$

Lemma A.3. *Assume that conditions* C1–C5 *hold. For any integer $r \geq 2$, we have, uniformly over $i = 1,\dots,n$, $j = 1,\dots,m_i$,*

$$E\{|\hat{\varphi}(X_{ij}^{\mathrm{T}}\beta;\beta,\theta)-\varphi(X_{ij}^{\mathrm{T}}\beta)|^r\} = O(h^{2r}) + O(n^{-r/2}h^{1-r}), \tag{A.8}$$

$$E\{|\hat{g}'(X_{ij}^{\mathrm{T}}\beta;\beta,\theta)-g'(X_{ij}^{\mathrm{T}}\beta)|^r\} = O(h^{r}) + O(n^{-r/2}h^{1-2r}), \tag{A.9}$$

where $\hat{\varphi}(\cdot) = \hat{g}(\cdot),\ \hat{g}_{1s}(\cdot)$ and $\hat{g}_{2k}(\cdot),\ \varphi(\cdot) = g(\cdot),\ g_{1s}(\cdot)$ and $g_{2k}(\cdot),\ 1 \le s \le p,\ 1 \le k \le q$.

Lemma A.4. *Under the assumptions of Theorem 2.1, if $(\beta^{(r)},\theta)$ is the true value of the parameter, and the rth component of β is a positive number, then*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\eta}_i(\beta^{(r)},\theta) \xrightarrow{L} N(0,\Omega(\beta^{(r)},\theta)),$$

where $\Omega(\beta^{(r)},\theta) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\{[\Lambda_i - E(\Lambda_i|X_i\beta)]V_i^{-1}(e_ie_i^{\mathrm{T}})V_i^{-1}[\Lambda_i - E(\Lambda_i|X_i\beta)]^{\mathrm{T}}\}$ and Λ_i is defined in (2.1).

Proof. In matrix notation defined in Section 2, it is easy to show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\eta}_i(\beta^{(r)},\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\Lambda_i - E(\Lambda_i|X_i\beta)]V_i^{-1}e_i + \sum_{\nu=1}^{4}M_\nu, \tag{A.10}$$

where

$$
\begin{aligned}
M_1 &= (M_{11}^{\mathrm{T}} J_{\beta^{(r)}}, M_{12}^{\mathrm{T}})^{\mathrm{T}}, \qquad M_2 = (M_{21}^{\mathrm{T}} J_{\beta^{(r)}}, \mathbf{0})^{\mathrm{T}},\\
M_3 &= (M_{31}^{\mathrm{T}} J_{\beta^{(r)}}, M_{32}^{\mathrm{T}})^{\mathrm{T}}, \qquad M_4 = (M_{41}^{\mathrm{T}} J_{\beta^{(r)}}, \mathbf{0})^{\mathrm{T}},\\
M_{11} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{[\hat{G}'_{\Delta}(X_i\beta; \beta, \theta) - G'_{\Delta}(X_i\beta)][X_i - \hat{G}_1(X_i\beta; \beta)]\}^{\mathrm{T}} V_i^{-1} e_i,\\
M_{12} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [G_2(X_i\beta) - \hat{G}_2(X_i\beta; \beta)]^{\mathrm{T}} V_i^{-1} e_i,\\
M_{21} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{G'_{\Delta}(X_i\beta)[G_1(X_i\beta) - \hat{G}_1(X_i\beta; \beta)]\}^{\mathrm{T}} V_i^{-1} e_i,\\
M_{31} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{G'_{\Delta}(X_i\beta)[X_i - \hat{G}_1(X_i\beta; \beta)]\}^{\mathrm{T}} V_i^{-1} [G(X_i\beta) - \hat{G}(X_i\beta; \beta, \theta)],\\
M_{32} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Z_i - \hat{G}_2(X_i\beta; \beta)]^{\mathrm{T}} V_i^{-1} [G(X_i\beta) - \hat{G}(X_i\beta; \beta, \theta)],\\
M_{41} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{[\hat{G}'_{\Delta}(X_i\beta; \beta, \theta) - G'_{\Delta}(X_i\beta)][X_i - \hat{G}_1(X_i\beta; \beta)]\}^{\mathrm{T}} V_i^{-1} [G(X_i\beta) - \hat{G}(X_i\beta; \beta, \theta)].
\end{aligned}
$$

To prove Lemma A.4, we first need to prove that

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\Lambda_i - E(\Lambda_i | X_i\beta)] V_i^{-1} e_i \xrightarrow{L} N(0, \Omega(\beta^{(r)}, \theta)). \tag{A.11}
$$

For the convenience of proof, let

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\Lambda_i - E(\Lambda_i | X_i\beta)] V_i^{-1} e_i \hat{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{in}.
$$

By $E(e_i|X_i) = 0$ and condition C4, and $\mathrm{Cov}(\xi_{in}) = E\{[\Lambda_i - E(\Lambda_i|X_i\beta)] V_i^{-1} (e_i e_i^{\mathrm{T}}) V_i^{-1} [\Lambda_i - E(\Lambda_i|X_i\beta)]^{\mathrm{T}}\}$. Let

$$
\Omega(\beta^{(r)}, \theta) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathrm{Cov}(\xi_{in}),
$$

it follows that the sequence of kth elements $\{\xi_{in}^{(k)}\}$ of $\{\xi_{in}\}$ $(k = 1, \ldots, p+q-1)$ satisfy, for any given $\varepsilon > 0$, $\frac{1}{n}\sum_{i=1}^{n} E\{\xi_{in}^{(k)^2} I(|\xi_{in}^{(k)}| > \varepsilon n^{1/2})\} \to 0$ $(n \to \infty)$. This means that the Lindeberg condition for the central limit theorem holds. Therefore, (A.11) follows.

To prove Lemma A.4, we only need to show that $M_{l1} \xrightarrow{P} 0$ $(l = 1, 2, 3, 4)$, $M_{12} \xrightarrow{P} 0$, $M_{32} \xrightarrow{P} 0$. Consider M_{11}. Let $M_{11,s}$ denote the sth component of M_{11} $(1 \le s \le p)$. Let $\widetilde{X}_{ijs} = X_{ijs} - \hat{g}_{1s}(X_{ij}^{\mathrm{T}}\beta; \beta)$, and σ_i^{jl} be the (j, l)th element of V_i^{-1} $(i = 1, \ldots, n, j = 1, \ldots, m_i, l = 1, \ldots, m_i)$. Then, we have

$$
\begin{aligned}
M_{11,s} = {} & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{l=1}^{m_i} \left[\sum_{i_1=1}^{n} \sum_{j_1=1}^{m_{i_1}} \widetilde{W}_{ni_1j_1}(X_{ij}^{\mathrm{T}}\beta; \beta) g(X_{i_1j_1}^{\mathrm{T}}\beta) - g'(X_{ij}^{\mathrm{T}}\beta) \right] \widetilde{X}_{ijs} \sigma_i^{jl} e_{il}\\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \widetilde{W}_{nij}(X_{ij}^{\mathrm{T}}\beta; \beta) \widetilde{X}_{ijs} \sigma_i^{jj} e_{ij}^2 + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j\neq l}^{m_i} \widetilde{W}_{nil}(X_{ij}^{\mathrm{T}}\beta; \beta) \widetilde{X}_{ijs} \sigma_i^{jl} e_{il}^2\\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{l=1}^{m_i} \sum_{i_1\neq i}^{n} \sum_{j_1\neq l}^{m_{i_1}} \widetilde{W}_{ni_1j_1}(X_{ij}^{\mathrm{T}}\beta; \beta) \widetilde{X}_{ijs} \sigma_i^{jl} e_{il} e_{i_1j_1}\\
\hat{=} {} & M_{11,s}^{(1)} + M_{11,s}^{(2)} + M_{11,s}^{(3)} + M_{11,s}^{(4)}.
\end{aligned}
$$

By condition C1 and Lemma A.3, we can prove that $E(\widetilde{X}_{ijs}^4) < \infty$. From conditions C4 and C7, and Lemma A.1, and using the Cauchy–Schwarz inequality and invoking the Lemma 4.2 in [11], we have

$$\begin{aligned} E(M_{11,s}^{(1)^2}) &\le cE\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j,l=1}^{m_i}\left[\sum_{i_1=1}^{n}\sum_{j_1=1}^{m_{i_1}}\widetilde{W}_{ni_1j_1}(X_{ij}^{\mathrm{T}}\beta;\beta)g(X_{i_1j_1}^{\mathrm{T}}\beta)-g'(X_{ij}^{\mathrm{T}}\beta)\right]^2\widetilde{X}_{ijs}^2\sigma_i^{jl^2}E(e_{il}^2|X_{ijs})\right\} \\ &\le cn^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}E^{1/2}\left[\sum_{i_1=1}^{n}\sum_{j_1=1}^{m_{i_1}}\widetilde{W}_{ni_1j_1}(X_{ij}^{\mathrm{T}}\beta;\beta)g(X_{i_1j_1}^{\mathrm{T}}\beta)-g'(X_{ij}^{\mathrm{T}}\beta)\right]^4E^{1/2}(\widetilde{X}_{ijs}^4) \\ &\le ch^2 \to 0. \end{aligned}$$

Invoking the conditions C4 and C7, Lemma A.2 and $E(\widetilde{X}_{ijs}^4) < \infty$, we have

$$\begin{aligned} E(|M_{11,s}^{(2)}|) &\le cn^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}E^{1/4}\{\widetilde{W}_{nij}^4(X_{ij}^{\mathrm{T}}\beta;\beta)\}E^{1/4}(\widetilde{X}_{ijs}^4)E^{1/2}(e_{ij}^4|\widetilde{X}_{ijs}) \\ &\le c(nh^2)^{-1/2}+c(nh^{5/2})^{-1}\to 0. \end{aligned}$$

Similarly, we can prove that $E(|M_{11,s}^{(3)}|) \to 0$. For $M_{11,s}^{(4)}$, let $a_{ij,i_1j_1} = \widetilde{W}_{ni_1j_1}(X_{ij}^{\mathrm{T}}\beta;\beta)\widetilde{X}_{ijs} + \widetilde{W}_{nij}(X_{i_1j_1}^{\mathrm{T}}\beta;\beta)\widetilde{X}_{ijs}$ and data set $\mathcal{X} = \{X_{ij}, i = 1,\ldots,n, j = 1,\ldots,m_i\}$. Note that $\sup_x E\{e_{ij}^2|X_{ijs} = x\} \le c < \infty$ and the independence of e_{ij} among the different subjects given the data set $\mathcal{X}$, again using Lemma A.2, we have

$$\begin{aligned} E(M_{11,s}^{(4)2}) &= E\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{l=1}^{m_i}\sum_{i_1\ne i}^{n}\sum_{j_1\ne l}^{m_{i_1}}\widetilde{W}_{ni_1j_1}(X_{ij}^{\mathrm{T}}\beta;\beta)\widetilde{X}_{ijs}\sigma_i^{jl}e_{il}e_{i_1j_1}\right\}^2 \\ &= E\left\{\frac{1}{\sqrt{n}}\sum_{i=2}^{n}\sum_{j=1}^{m_i}\sum_{l=1}^{m_i}\sum_{i_1=1}^{i-1}\sum_{j_1=1}^{m_{i_1}}a_{ij,i_1j_1}\sigma_i^{jl}e_{il}e_{i_1j_1}\right\}^2 \\ &= n^{-1}\sum_{i=2}^{n}\sum_{j=1}^{m_i}\sum_{l=1}^{m_i}\sum_{i_1=1}^{i-1}\sum_{j_1=1}^{m_{i_1}}E\{a_{ij,i_1j_1}^2\sigma_i^{jl^2}E(e_{il}^2|\widetilde{X}_{ijs})E(e_{i_1j_1}^2|\widetilde{X}_{ijs})\} \\ &\le cn^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{i_1\ne i}^{n}\sum_{j_1\ne j}^{m_{i_1}}E^{1/2}\{\widetilde{W}_{ni_1j_1}^4(X_{ij}^{\mathrm{T}}\beta;\beta)\}E^{1/2}(\widetilde{X}_{ijs}^4) \\ &\le c(nh^2)^{-1}+c(nh^{5/2})^{-2}\to 0. \end{aligned}$$

Consequently, the moment of sth component of M_{11} converges to 0. By the Markov inequality, we can prove that $M_{11,s} \xrightarrow{P} 0$ for each $1 \le s \le p$. Therefore, we obtain that $M_{11} \xrightarrow{P} 0$. Let $M_{21,s}$ denote the sth component of M_{21}. Then

$$M_{21,s} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{l=1}^{m_i}g'(X_{ij}^{\mathrm{T}}\beta)[g_{1s}(X_{ij}^{\mathrm{T}}\beta)-\hat{g}_{1s}(X_{ij}^{\mathrm{T}}\beta;\beta)]\sigma_i^{jl}e_{il}.$$

From Lemma A.3, and conditions C4 and C7, we have

$$\begin{aligned} E(M_{21,s}^2) &\le cn^{-1}E\left[\sum_{i=1}^{n}\sum_{j=1}^{m_i}g'(X_{ij}^{\mathrm{T}}\beta)^2\{g_{1s}(X_{ij}^{\mathrm{T}}\beta)-\hat{g}_{1s}(X_{ij}^{\mathrm{T}}\beta;\beta)\}^2\right] \\ &\le ch^4+c(nh)^{-1}\to 0. \end{aligned}$$

This yields $M_{21} \xrightarrow{P} 0$. Similarly, we obtain that $M_{12} \xrightarrow{P} 0$. Hence, we have proved that

$$M_i \xrightarrow{P} 0, \quad i = 1, 2. \tag{A.12}$$

Similar arguments to the proofs of (A.12), we can also obtain that $M_3 \xrightarrow{P} 0$ and $M_4 \xrightarrow{P} 0$. This, together with (A.10)–(A.12), proves Lemma A.4. □

Lemma A.5. *Under the assumptions of Theorem 2.1, if $(\beta^{(r)}, \theta)$ is the true value of the parameter, and the rth component of β is a positive number, we have*

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_i(\beta^{(r)},\theta)\hat{\eta}_i^{\mathrm{T}}(\beta^{(r)},\theta) \xrightarrow{P} \Omega(\beta^{(r)},\theta),$$

where $\hat{\eta}_i(\beta^{(r)},\theta)$ is defined in (2.9), $\Omega(\beta^{(r)},\theta)$ is defined in condition C9 .

Proof. By Lemmas A.1–A.4, the proof of Lemma A.5 is similar to that of Lemma 5 in [14], we then omit the details here. □

Lemma A.6. *Under the assumptions of Theorem 2.1, if $(\beta^{(r)}, \theta)$ is the true value of the parameter, and the rth component of β is a positive number, we have*

$$\max_{1\le i\le n} \|\hat{\eta}_i(\beta^{(r)}, \theta)\| = o_P(n^{1/2}).$$

Proof. By Lemmas A.1–A.5, we can prove Lemma A.6 by using the same arguments used in the proof of Lemma 6 in [13]. □

Lemma A.7. *Under the assumptions of Lemma A.4, we have*

$$\|\lambda\| = O_P(n^{-1/2}).$$

Proof. Let $\lambda = \gamma\phi$, where $\gamma \ge 0$, $\phi \in R^{p+q-1}$ and $\|\phi\| = 1$. By (2.14), we have

$$\begin{aligned}
0 &= \frac{1}{n}\phi^{\mathrm{T}} \sum_{i=1}^{n} \frac{\hat{\eta}_i(\beta^{(r)}, \theta)}{1+\gamma\phi^{\mathrm{T}}\hat{\eta}_i(\beta^{(r)}, \theta)} \\
&= \frac{1}{n}\phi^{\mathrm{T}} \sum_{i=1}^{n} \hat{\eta}_i(\beta^{(r)}, \theta) - \frac{1}{n}\gamma \sum_{i=1}^{n} \frac{(\phi^{\mathrm{T}}\hat{\eta}_i(\beta^{(r)}, \theta))^2}{1+\gamma\phi^{\mathrm{T}}\hat{\eta}_i(\beta^{(r)}, \theta)} \\
&\le \left|\frac{1}{n}\phi^{\mathrm{T}} \sum_{i=1}^{n} \hat{\eta}_i(\beta^{(r)}, \theta)\right| - \frac{\gamma}{1+\gamma \max\limits_{1\le i\le n}\|\hat{\eta}_i(\beta^{(r)}, \theta)\|} \frac{1}{n}\sum_{i=1}^{n}(\phi^{\mathrm{T}}\hat{\eta}_i(\beta^{(r)}, \theta))^2 \\
&\le \left|\frac{1}{n}\phi^{\mathrm{T}} \sum_{i=1}^{n} \hat{\eta}_i(\beta^{(r)}, \theta)\right| - \frac{\gamma}{1+\gamma \max\limits_{1\le i\le n}\|\hat{\eta}_i(\beta^{(r)}, \theta)\|} \mathrm{mineig}(S(\beta^{(r)}, \theta)),
\end{aligned}$$

where $S(\beta^{(r)}, \theta) = \frac{1}{n}\sum_{i=1}^{n} \hat{\eta}_i(\beta^{(r)}, \theta)\hat{\eta}_i^{\mathrm{T}}(\beta^{(r)}, \theta)$ and mineig(A) denotes the minimum eigenvalue of matrix A. Therefore, we have

$$\gamma\,\mathrm{mineig}(S(\beta^{(r)}, \theta)) - \gamma\frac{1}{n}\phi^{\mathrm{T}}\sum_{i=1}^{n}\hat{\eta}_i(\beta^{(r)}, \theta)\max_{1\le i\le n}\|\hat{\eta}_i(\beta^{(r)}, \theta)\| \le \left|\frac{1}{n}\phi^{\mathrm{T}}\sum_{i=1}^{n}\hat{\eta}_i(\beta^{(r)}, \theta)\right|. \tag{A.13}$$

By Lemmas A.4 and A.5, we have

$$\frac{1}{n}\phi^{\mathrm{T}}\sum_{i=1}^{n}\hat{\eta}_i(\beta^{(r)}, \theta) = O_P(n^{-1/2}), \qquad \mathrm{mineig}(S(\beta^{(r)}, \theta)) = O_P(1). \tag{A.14}$$

By Lemma A.6, (A.13) and (A.14), $\gamma = O_P(n^{1/2})$, that is, $\|\lambda\| = O_P(n^{1/2})$. □

References

[1] J.M. Chiou, H.G. Müller, Estimated estimating equations: Semiparametric inference for clustered and longitudinal data, J. Roy. Statist. Soc. B 67 (4) (2005) 531–553.
[2] S.L. Zeger, P.J. Diggle, Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters, Biometrics 50 (1994) 689–699.
[3] X.H. Lin, R.J. Carrol, Semiparametric regression for clustered data using generalized estimating equations, J. Amer. Statist. Assoc. 96 (2001) 1045–1056.
[4] X.M. He, Z.Y. Zhu, W.K. Fung, Estimation in a semiparametric model for longitudinal data with unspecified dependence structure, Biometrika 89 (3) (2002) 579–590.
[5] J.Q. Fan, R.Z. Li, New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis, J. Amer. Statist. Assoc. 99 (2004) 710–723.
[6] W.M. Qian, Method of estimation in statistical models for longitudinal data. Ph.D. Thesis, Tongji University, Shanghai, 2003.
[7] X.Q. Sun, J.H. You, Iterative weighted partial spline least squares estimation in semiparametric modeling of longitudinal data, Sci. China, Ser. A 46 (5) (2003) 724–735.
[8] J.H. You, G.M. Chen, Y. Zhou, Block empirical likelihood for longitudinal partially linear regression models, Canadian J. Statist. 34 (1) (2006) 79–96.
[9] L.G. Xue, L.X. Zhu, Empirical likelihood semiparametric regression analysis for longitudinal data, Biometrika 94 (2007) 921–937.
[10] J.Q. Fan, T. Huang, R.Z. Li, Analysis of longitudinal data with semiparametric estimation of covariance function, J. Amer. Statist. Assoc. 102 (2007) 632–641.
[11] G.R. Li, P. Tian, L.G. Xue, Generalized empirical likelihood inference in semiparametric regression model for longitudinal data, Acta Math. Sin. (Engl. Ser.) 24 (12) (2008) 2029–2040.
[12] R.J. Carroll, J.Q. Fan, I. Gijbels, M.P. Wand, Generalized partially linear single-index models, J. Amer. Statist. Assoc. 92 (1997) 477–489.
[13] L.G. Xue, L.X. Zhu, Empirical likelihood confidence regions of the parameters in a partially linear single-index model, Sci. China, Ser. A 48 (10) (2005) 1333–1348.
[14] L.X. Zhu, L.G. Xue, Empirical likelihood confidence regions in a partially linear single-index model, J. Roy. Statist. Soc. B 68 (2006) 549–570.
[15] Y. Yu, D. Ruppert, Penalized spline estimation for partially linear single-index models, J. Amer. Statist. Assoc. 97 (2002) 1042–1054.
[16] Y. Xia, W. Härdle, Semi-parametric estimation of partially linear single-index models, J. Multivariate Anal. 97 (2006) 1162–1184.

[17] A.B. Owen, Empirical likelihood ratio confidence intervals for a single function, Biometrika 75 (2) (1988) 237–249.
[18] A.B. Owen, Empirical likelihood ratio confidence regions, Ann. Statist. 18 (1990) 90–120.
[19] A.B. Owen, Empirical Likelihood, Chapman & Hall/CRC, New York, 2001.
[20] Y. Kitamura, Empirical likelihood methods with weakly dependent processes, Ann. Statist. 25 (1997) 2084–2102.
[21] L.G. Xue, L.X. Zhu, Empirical likelihood for a varying coefficient model with longitudinal data, J. Amer. Statist. Assoc. 102 (2007) 642–654.
[22] K.Y. Liang, S.L. Zeger, Longitudinal data analysis using generalized linear model, Biometrika 73 (1986) 13–22.
[23] J.Q. Fan, I. Gijbels, Local Polynomial Modeling and Its Applications, Chapman and Hall, London, 1996.
[24] X.H. Lin, R.J. Carrol, Nonparametric function estimation for clustered data when the predictor is measured without/with error, J. Amer. Statist. Assoc. 95 (2000) 520–534.
[25] L.G. Xue, L.X. Zhu, Empirical likelihood for single-index models, J. Multivariate Anal. 97 (6) (2006) 1295–1312.
[26] K.C. Li, Slice inverse regression for dimension reduction, J. Amer. Statist. Assoc. 86 (1991) 316–327.
[27] R.A. Kaslow, D.G. Ostrow, R. Detels, J.P. Phair, B.F. Polk, C.R. Rinaldo, The multicenter AIDS cohort study: Rationale, organization and selected characteristics of the participants, Am. J. Epidemiol. 126 (1987) 310–318.
[28] C.O. Wu, C.T. Chiang, D.R. Hoover, Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data, J. Amer. Statist. Assoc. 93 (1998) 1388–1402.
[29] J.Z. Huang, C.O. Wu, L. Zhou, Varying-coefficient models and basis function approximations for the analysis of repeated measurements, Biometrika 89 (2002) 111–128.
[30] J.Q. Fan, J. Zhang, Two-step estimation of functional linear models with applications to longitudinal data, J. Roy. Statist. Soc. B 62 (2000) 303–322.